

REMARKS

This communication is in response to the Office Action mailed on September 13, 2007, a telephone interview conducted on February 5, 2008, and receipt of an interview summary and advisory action both mailed from the patent office on or around February 7, 2008. The examiner is thanked for his time and helpful suggestions. In the Office Action claims 1-31 were pending of which claims 1-31 were rejected.

Rejections based on 35 U.S.C. §101

As discussed during the interview, applicants hereby disclaim any prior argument regarding communication media as being computer readable storage media. Claims 1-13 each recite computer readable storage media which as stated by the examiner is statutory. Withdrawal of the above rejection is respectfully requested.

Information disclosure statement

The Office Action reports that the Information Disclosure Statement (IDS) filed on 8/17/07 failed to comply with 37 C.F.R. 1.98(a)(3) because it did not include a concise statement of relevance as understood by the individual designated in 37 C.F.R. 1.56(c). It is believed that the IDS includes a statement of relevance, "This paper is an overview of the state-of-the-art of methods for the Chinese word segmentation task, in particular some investigations of overlapping ambiguity distribution in the corpus, and the overlapping ambiguity detection coverage of the FMM+BMM method. It is asserted that this statement is from the inventor, a Chinese citizen, and is a concise statement of relevance as understood by this individual. Thus, it is believed that the IDS meets the requirements of 37 C.F.R. 1.98(a)(3). However, it is also believed that a complete English translation of the Chinese language reference is not available at this time.

Rejections based on obviousness

The Office Action next reports that claims 1-4, 6-7, 14-21, 23, 25-26, and 28 were rejected under 35 U.S.C. §103 as being unpatentable over U.S. Patent No. 5,806,021 to Chen et al. (hereinafter "Chen") in view of U.S. Patent No. 6,968,308 to Brockett et al. (hereinafter "Brockett") It is respectfully submitted that the cited references even when combined do not teach or suggest all of the features of claim 1.

During the interview, the applicants' attorneys explained an important distinction between segmented languages such as English and unsegmented languages such as Chinese. Segmented languages are segmented in that the boundaries between individual words are readily apparent, typically by spaces between words. In contrast, in an unsegmented language the boundaries between individual words is not readily apparent. For example, in Chinese, a Chinese word can include one, two, or more consecutive Chinese characters arranged in a string. Thus, during natural language processing it is typical to process text of an unsegmented language by segmenting the text into individual words before further processing takes place. However, sometimes the location of a word boundary is ambiguous. As discussed during the interview an illustration of this type of ambiguity is as follows: suppose "ABC" is a string of characters "A", "B", and "C" of an unsegmented language. The string "ABC" is an overlapping ambiguity string if "AB" and "BC" are both recognizable words in the unsegmented language. Thus, possible segmentations of string "ABC" include "AB/C" and "A/BC". As explained in the Specification, overlapping ambiguity strings often cause segmentation errors. [See Specification, page 3, lines 21-28]

Without conceding that the prior art reads on the previously presented claims and thus reserving the right to file a continuation based on the previously presented claims, claim 1 has been amended to recite, a computer readable storage media storing

instructions readable by a computer which, when implemented, cause the computer perform steps comprising segmenting a sentence of Chinese characters into constituent Chinese words having one or more Chinese characters; recognizing an overlapping ambiguity string in the segmented sentence, wherein the overlapping ambiguity string comprises at least three Chinese characters having at least two possible segmentations; obtaining probability information based on at least one context feature adjacent the overlapping ambiguity string; and outputting an indication for selecting one of the at least two possible segmentations as a function of the obtained probability information. [emphasis added]

Thus, claim 1 has been amended to clarify that the unsegmented language is Chinese and to further clarify that an overlapping ambiguity string is a string of at least three Chinese characters that have at least two possible segmentations. After the overlapping ambiguity string is recognized, probability information is obtained that is based on a least one context feature (e.g. grammatical symbols or other Chinese characters) adjacent or surrounding the overlapping ambiguity string. Then an indication for selecting one of the at least two possible segmentations is output.

It is respectfully submitted that the cited references, alone or in combination, do not teach, suggest or render obvious all of the features of claim 1. The primary reference Chen discloses a segmenter that performs continuous segmentation of text using at least two approaches. The segmenter can use Forward-Backward Maximum Matching where the segmentation (either forward or backward) is selected based on the likelihood. The Chen segmenter can also use a statistical stack method which is slower than the Forward-Backward Maximum Matching method if accuracy and not speed is the primary concern. [Abstract]

However, it is believed that Chen is directed towards continuous segmentation of text and not primarily to resolving

overlapping ambiguity strings in Chinese as are the present claims. Further, it is believed that Chen nowhere discloses obtaining probability information based on at least one context feature adjacent or surrounding the overlapping ambiguity string and the selecting one of at least two possible segmentations based on the obtained probability information as recited in claim 1.

The Office Action does admit that Chen does not disclose recognizing the overlapping ambiguity string in the input sentence as a function of the two segmentations, obtaining probability information based on at least one context feature adjacent the overlapping ambiguity string and at least part of the recognized OAS for each of the FMM and BMM, outputting an indication for selecting one of the two segmentations as a function of the obtained probability information, and replacing the overlapping ambiguity string with tokens. It is submitted that these missing features are important features of claim 1. Further, it is noted that claim 1 has been amended with additional features. Thus, it is believed that Chen does not teach or suggest all the features of claim 1.

The secondary reference Brockett discloses a method of segmenting non-segmented text using syntactic parse. However, Brockett is related to the Japanese language which uses four different kinds of script including Kanji, haragana, katakana, and roma. These four scripts can be used to spell the same word. Thus, the methods of segmenting according to Brockett are designed to segment words in a way that accounts for these variations. [Col. 2, lines 4-8] In contrast, as discussed in the interview, claim 1 has been amended to be limited to Chinese language. Thus, it is believed that Brockett does not teach or suggest all the features of claim 1.

In view of the foregoing, it is believed that claim 1 is patentable over the cited art. Claims 2-12 depend on claim 1 and are believed to be separately patentable. Reconsideration and

allowance of claims 1-12 are respectfully requested. Claim 12 has been cancelled with its features being added to independent claim 1.

Claim 14 was also rejected based on the same combination of Chen and Brockett. As stated above, without admitting that the cited combination reads on the previously presented claims, claim 14 has been amended to recite a method of segmentation of a sentence of Chinese text, the sentence having an overlapping ambiguity string, the method comprising generating a Forward Maximum Matching (FMM) segmentation of the sentence; generating a Backward Maximum Matching (BMM) segmentation of the sentence; recognizing the overlapping ambiguity string based on a difference between the FMM segmentation and the BMM segmentation; obtaining probability information based on at least one context feature surrounding the overlapping ambiguity string and at least part of the overlapping ambiguity string; and outputting an indication for selecting one of the FMM segmentation and the BMM segmentation as a function of obtained probability information. [emphasis added]

Claim 14 has also been amended in a manner similar to claim 1. Thus, the remarks above are hereby incorporated by reference. Claim 14 now clarifies that the unsegmented language is Chinese. Claim 14 further specifies that the overlapping ambiguity string is recognized based on a difference between the FMM segmentation and the BMM segmentation. Also, claim 14 clarifies that the probability information that is obtained is based on one or more context features and at least part of the overlapping ambiguity string. It is believed that the cited combination, singly or in combination, does not teach or suggest all of the features of claim 14. Thus, claim 14 is believed to be patentable over the cited art. Claims 15-24 depend on claim 14 and are believed to be separately patentable. Reconsideration and allowance of claims 14-24 are respectfully requested.

The Office Action further cites the same combination against independent claim 25. Claim 25 has been amended to recite a method of segmenting a sentence of Chinese text comprising: recognizing an overlapping ambiguity string in the sentence; receiving probability information from an N-gram language model comprising probability information for constituent words of the overlapping ambiguity string and context features surrounding the overlapping ambiguity string; and resolving the overlapping ambiguity string based on the received probability information. [emphasis added]

The discussion of the cited references is hereby incorporated by reference. Claim 25 has been amended so that the a sentence of Chinese language is segmented, an overlapping ambiguity string is recognized in the sentence, N-gram probability information for constituent words of the overlapping ambiguity string and context features surrounding the overlapping ambiguity string is received, and the overlapping ambiguity string is resolved based on this received probability information. As discussed above, it is submitted that the cited references do not teach or suggest all of the features of claim 25.

In light of the foregoing, it is believed that claim 25 is patentable over the cited art. Claims 26-31 depend on claim 25 and are believed to be separately patentable. Reconsideration and allowance of claims 25-31 is respectfully requested.

The foregoing remarks are intended to assist the Office in examining the application and in the course of explanation may employ shortened or more specific or variant descriptions of some of the claim language. Such descriptions are not intended to limit the scope of the claims; the actual claim language should be considered in each case. Furthermore, the remarks are not to be considered exhaustive of the facets of the invention which are rendered patentable, being only examples of certain advantageous features and differences, which applicant's attorney chooses to

mention at this time. For the foregoing reasons, applicant reserves the right to submit additional evidence showing the distinction between applicant's invention to be unobvious in view of the prior art.

Furthermore, in commenting on the references and in order to facilitate a better understanding of the differences that are expressed in the claims, certain details of distinction between the same and the present invention have been mentioned, even though such differences do not appear in all of the claims. It is not intended by mentioning any such unclaimed distinctions to create any implied limitations in the claims.

The Director is authorized to charge any fee deficiency required by this paper or credit any overpayment to Deposit Account No. 23-1123.

Respectfully submitted,

WESTMAN, CHAMPLIN & KELLY, P.A.

By: /Linda Ji/

Linda Ji, Reg. No. 49,027  
900 Second Avenue South, Suite 1400  
Minneapolis, Minnesota 55402-3319  
Phone: (612) 334-3222 Fax: (612) 334-3312